Pitfalls in survey measurements of economic parameters

Pitfalls in survey measurements of economic parameters

P J Thomas

Professor of Risk Management, Safety Systems Research Centre, South West Nuclear Hub, University of Bristol, Queen's Building, University Walk, , Bristol BS8 1TR , UK

E-mail: philip.thomas@bristol.ac.uk

Abstract. Opinion surveys are sometimes used in attempts to measure important economic parameters in the absence of a free market. It is then important that the most rigorous standards are applied if the survey results are to be meaningful. The paper highlights new theoretical results in this field and documents examples where lack of rigour has led to serious consequences.

1. Introduction: fundamental rules for interpreting opinion surveys

There is a generally accepted hierarchy of methods [1] to be used in measuring the value of any good, in order of preference:

1. the market value if a free market in the good exists – the best way
2. the value deduced from revealed preferences – "I have always thought the actions of men the best interpreters of their thoughts" – John Locke [2]
3. the value deduced from stated preferences, typically from opinion surveys – "respondents in stated preference surveys may have an incentive to deliberately misrepresent their true preferences in order to achieve a more desirable outcome for themselves ... individuals may overstate their valuations of the good if they believe their responses influence its provision and are unrelated to the price they will be charged for it" [3].

The lack of a free market makes it hard to place a value on public goods such as the amenity of access to the countryside or of keeping pollutants out of the world's oceans or of the continued existence of an endangered species. Revealed preference techniques may require considerable economic insight, so opinion surveys are sometimes used as a last resort. But high standards need to be applied if this last process is to be meaningful. Recent papers [4], [5] have highlighted some constraints of which any opinion pollster needs to be aware.

Opinion surveys aim to consolidate different people's judgements of the value of a numerical parameter into a single figure representative of the population as a whole. Clearly the sample size should not be too small, but the well-known concept of "margin of error" gives useful guidance only for the restricted case of two possible judgements, when, for example, the views of at least 1,067 people are needed to give a margin of error below 3% [6]. The sample size may need to be higher when the parameter under investigation is continuous rather binary.

Once the size of the sample has been chosen, it is necessary to decide on the correct consolidation statistic. "Structural view independence" (SVI) [4] is a new criterion that requires the view of each person in the sample to be accorded the same importance. The consolidation process should not be structurally dependent on the size of the view, which builds in a bias the direction of which may be chosen by the surveyor. Lack of SVI will mean that each person's view is filtered by that of at least one other person. Many central measures employed in the past (e.g. geometric mean and root-mean-square) have failed the SVI test and should not be used. Censoring (or "trimming") is clearly in violation as it gives zero weight to certain opinions, with the median, the ultimate in bilateral trimming, falling into this category. Only the arithmetic mean has been proved to satisfy SVI [4].

Some opinion pollsters have sought to test their main result with a sensitivity analysis, whereby the respondents are asked to place a value on a new parameter that is 3 times larger than the original. But in the frequent case where the original parameter will be seen by the respondents as a commodity, the 3 for the price of 2 pricing rule [5] will apply, when the new parameter bringing 3 times the benefit will be valued only twice as highly.

2. A warning from the finance sector: London Interbank Offered Rates (LIBORs)

LIBORs, benchmarks for interest rates globally, are referenced in transactions with a notional outstanding value of $300 trillion [7]. They determine the payments for both over-the-counter interest rate derivatives contracts and exchange-traded interest rate contracts, including by public authorities and small businesses. Their influence extends to mortgages and bank loans. In the absence of an active market in 2007 (banks were not lending to each other), a survey was instituted to establish the LIBOR. The performance of this 3rd-best option was made worse by rejecting the simple arithmetic average and censoring high and low values from the survey, with half the sample discarded (8 out of 16). Moreover each respondent had an incentive to falsify his view, as his bank would look more credit-worthy if he stated a low rate. The scene was set for one of the millennium's big banking scandals. Opinion censoring or "trimming" had a key role.

In September 2007, Bloomberg published an article identifying Barclays' LIBOR submissions as exceptionally high and questioning Barclays' credit-worthiness as a result. In March 2008, routine liquidity telephone calls to the Financial Services Authority revealed that "Manager D" felt his bank, Barclays, had been "picked upon for posting LIBORs above everyone else": "What is everybody, open brackets to be honest, including ourselves close brackets, going to do?", to which he provided the answer, "Keep their heads below the parapet and not stick out", and, a month later, "we did stick our heads above the parapet last year, got it shot off, and put it back down again. So, to the extent that, um, the LIBORs have been understated, are we guilty of being part of the pack? You could say we are. We've always been at the top end and therefore one of the four banks that's been eliminated".

The censoring inherent in the LIBOR survey method made such behaviour almost inevitable. Barclays was fined £290m, UBS £940m, RBS £390m, Icap £55m, Rabobank £660m, Lloyds Bank £226m and Deutsche Bank £1,700m. The Financial Conduct Authority now wishes to end LIBOR by 2021 because of its over-reliance on expert judgement in the absence of an active market [8].

3. Cautionary example of the UK "Value of a Prevented Fatality" (VPF)

The VPF still in use by UK Government Departments is based on a 1997 survey of 167 people. The surveyors though people had difficulty placing a value on a reduction in the risk of accidental death and devised a more complex, two-part process to find people's attitude to injuries and death. Based on the application of utility functions to model the satisfaction that an individual's wealth brings, the two-injury chained method [9] boils down to the following. The respondent is asked to imagine being injured in a road accident and being offered two hospital operations, A or B, each of which will leave him with a serious, "type 2" injury should it fail. Operation A will still leave the patient with a less serious, "type 1" injury even if it succeeds. But, a successful operation B will return the patient to full health almost immediately. Operation B looks to be the one to choose.

But what happens if the probabilities of success are different for the two surgical procedures? The surveyors thought that Operation B's potentially better outcome should cause the patient to accept a higher failure probability. They then theorised that information on acceptable failure probabilities, elicited via a "standard gamble" procedure, should determine the ratio of an individual's value of a prevented injury (VPI) of type 2 to his VPI for a type 1 injury.

Further theory from the surveyors suggested that an individual's VPI for a type 1 injury could be calculated directly after drawing out from the respondent both the maximum acceptable price (MAP, £) he would pay to avert the injury and the minimum acceptable compensation (MAC, £) he would be willing to receive to make up for the injury. Multiplying the individual's VPI for a type 1 injury by the VPI ratio just found then gives his VPI for a type 2 injury.

If the type 2 injury is taken to be fatal, the type 2 VPI will be equal to the individual's personal VPF. If the sample size is big enough to represent the full spectrum of wealths in the UK, then averaging over all the people in the sample will yield the VPF for the nation as a whole.

But using the surveyors' Constrained Power Utility function as the model for utility of wealth, the average wealth implied for the respondents came out as £1,730 under one injury ('W') and £5,252 under another ('X'). These figures (which should not be different, of course) represent 2% and 7% of the UK average of adult wealth (£78,300 at the time of the survey), so that obviously the UK population could not have been well represented [10] [11].

Worse than this, the two-injury chained method fails a fundamental test of its validity. Two methods of measuring the VPI for the same injury need to give the same value for each respondent ($y = x$) for the 2-injury chained method to be valid, but the best fit occurs for $y = 8.24x$, and even then the square of the correlation coefficient, R^2, is only 0.07. The surveyors claimed to be "fully aware of, and open about, this discrepancy" but decided to carry on anyway, seeing the method as "providing evidence which, when blended with judgement, helped consolidate the VPF" [12]: the new VPF came out only 10% different from the pre-existing figure. Judgement was made over 2 orders of magnitude, either through censoring or by swapping between the mean and the median as if they had equal validity (which is clearly not so).

Given the provenance just described, it would be surprising if the UK VPF were to turn out correct. In fact, taking the VPF to represent the valuation of population-average life expectancy in the UK, it is now clear that the VPF in current use is 4 times less than it should be. Its true value should be close to the equivalent figure used in the USA [14]. The recent validation of the objective J-value against pan-national data [15] has allowed this result to be calculated.

4. The UK VPF and three for the price of two

Prior to the survey discussed in Section 3, essentially the same group of investigators published the results of two earlier surveys aimed at determining people's willingness to pay to prevent a road fatality [16]. The first of these surveys involved 83 people and the second 52. The authors expected people's responses to be roughly linear in risk-reduction but in fact respondents seemed to be willing to pay only 1.4 times more for a three-fold reduction in risk in the first survey and only 1.3 times more in the second. The results of both these studies were rejected by the authors and their sponsors due to their perceived "serious doubt on the reliability and validity" of the study methods. The opinion pollsters went so far as to give their published paper the label, "Caveat Investigator" – "Let the investigator beware".

However the survey responses can be explained by the same reasoning that accounts for the prevalence of offers such as "buy one get one free" and "3 for the price of 2" in supermarkets – the theory of "Relative Utility Pricing" (RUP), which describes how consumers value a multi-pack of a commodity relative to a smaller pack [5]. A necessary feature in RUP theory is that a person should assign a positive utility to a positive benefit, with a larger benefit attracting a higher utility. This requirement on discrimination eliminated approximately 45% of the two samples from further analysis, leaving 74 respondents in total who were able to discriminate between the different safety measures or "packs". The mean ratios of the MAPs that these respondents were willing to pay were 1.98 in the first survey and 1.93 in the second. Recalling that the second "safety pack" provided three times the benefit of the first in both the surveys, the

MAP ratio of 2.0 deduced from RUP theory lies comfortably within the 90% confidence intervals of the observed means [17]. It is thus clear that the results of the first surveys should not have been dismissed in favour of the invalid survey method discussed in Section 4.

5. Conclusions

Opinion surveys have been used to measure parameters that have influenced and are influencing decisions involving life and death and trillions of pounds. It is thus indisputable that rigour in survey interpretation and hence measurement is essential. The examples discussed show how failing to apply the most rigorous measurement standards can have grave implications.

References

[1] PriceWaterhouseCoopers 2008 *Commentary US Financial Accounting Standards Board Statement FAS 157 – Fair Value Measurement* https://www.pwc.com/bm/en/publication/assets/usgaap_08_04.pdf

[2] Locke J 1690 *Essay Concerning Human Understanding* (London), http://www.earlymoderntexts.com/assets/pdfs/locke1690book1.pdf

[3] Fujiwara D and Campbell R 2011 *Valuation techniques for social cost-benefit analysis* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/209107/greenbook_valuation techniques.pdf

[4] Thomas P J 2014 " Structural view independence: A criterion for judging the objectivity of economic parameters measured by opinion survey" *Measurement*, Vol. 47 161–177.

[5] Thomas P and Chrystal A 2013 "Explaining the 'Buy One Get One Free' promotion: the Golden Ratio as a Marketing Tool" *American Journal of Industrial and Business Management* Vol. 3 No. 8 655-673.

[6] Larsen R J and Marx M L 2001 *An Introduction to Mathematical Statistics* 3rd Ed. (New Jersey: Prentice-Hall) p 330

[7] Thomas Philip 2013 "Libor lessons" *FT Adviser* 9 October https://www.ftadviser.com/2013/10/09/investments/economic-indicators/libor-lessons-oVUsJze4WmgSz1L8m763SK/article.html

[8] Financial Conduct Authority 2017 "The future of LIBOR" https://www.fca.org.uk/print/news/speeches/the-future-of-libor

[9] Carthy T Chilton S Covey J Hopkins L Jones-Lee M Loomes G Pidgeon N and Spencer A 1999 "On the contingent valuation of safety and the safety of contingent valuation: Part 2 – The CV/SG chained approach" *J. Risk Uncertain*. Vol. 17 187 – 213

[10] Thomas P J and Vaughan G J 2015 "Testing the validity of the 'value of a prevented fatality' (VPF) used to assess UK safety measures" *Process Safety and Environmental Protection* Vol 94 239 – 261.

[11] Thomas P J and Vaughan G J 2015 "Pitfalls in the application of utility functions to the valuation of human life" *Process Safety and Environmental Protection* Vol 98 148–169.

[12] Chilton S Covey J Hopkins L Jones-Lee M Loomes G Pidgeon N and Spencer A 2015 " Response to 'Testing the validity of the "value of a prevented fatality" (VPF) used to assess UK safety measures' " *Process Safety and Environmental Protection* Vol 93 293–298

[13] Thomas P J and Vaughan G J 2015 " 'Testing the validity of the "value of a prevented fatality" (VPF) used to assess UK safety measures': Reply to the comments of Chilton Covey Jones-Lee Loomes Pidgeon and Spencer" Process Safety and Environmental Protection Vol 93 pp 299 – 306

[14] U.S. Department of Transportation 2015 Guidance on Treatment of the Economic Value of a Statistical Life in U.S. Department of Transportation Analyses https://www transportation gov/regulations/economic-values-used-in-analysis

[15] Thomas P and Waddington I 2017 "Validating the J-value safety assessment tool against pan-national data" *Process Safety and Environmental Protection* Vol 112A 179 – 197

[16] Beattie J Covey J Dolan P Hopkins L Jones-Lee M Loomes G Pidgeon N Robinson A Spencer A 1998 "On the contingent valuation of safety and the safety of contingent valuation: Part 1 – Caveat investigator" *J Risk Uncertain* Vol 17 5 – 25

[17] Thomas P Vaughan G 2014 "Explaining Perceived Inconsistencies in 'Stated Preference' Valuations of Human Life" *American Journal of Industrial and Business Management* Vol 4 No 9 442 – 473.